TapNet: Neural Network Augmented with Task-Adaptive Projection for Few-Shot Learning

Sung Whan Yoon (presenter), Jun Seo and Jaekyun Moon

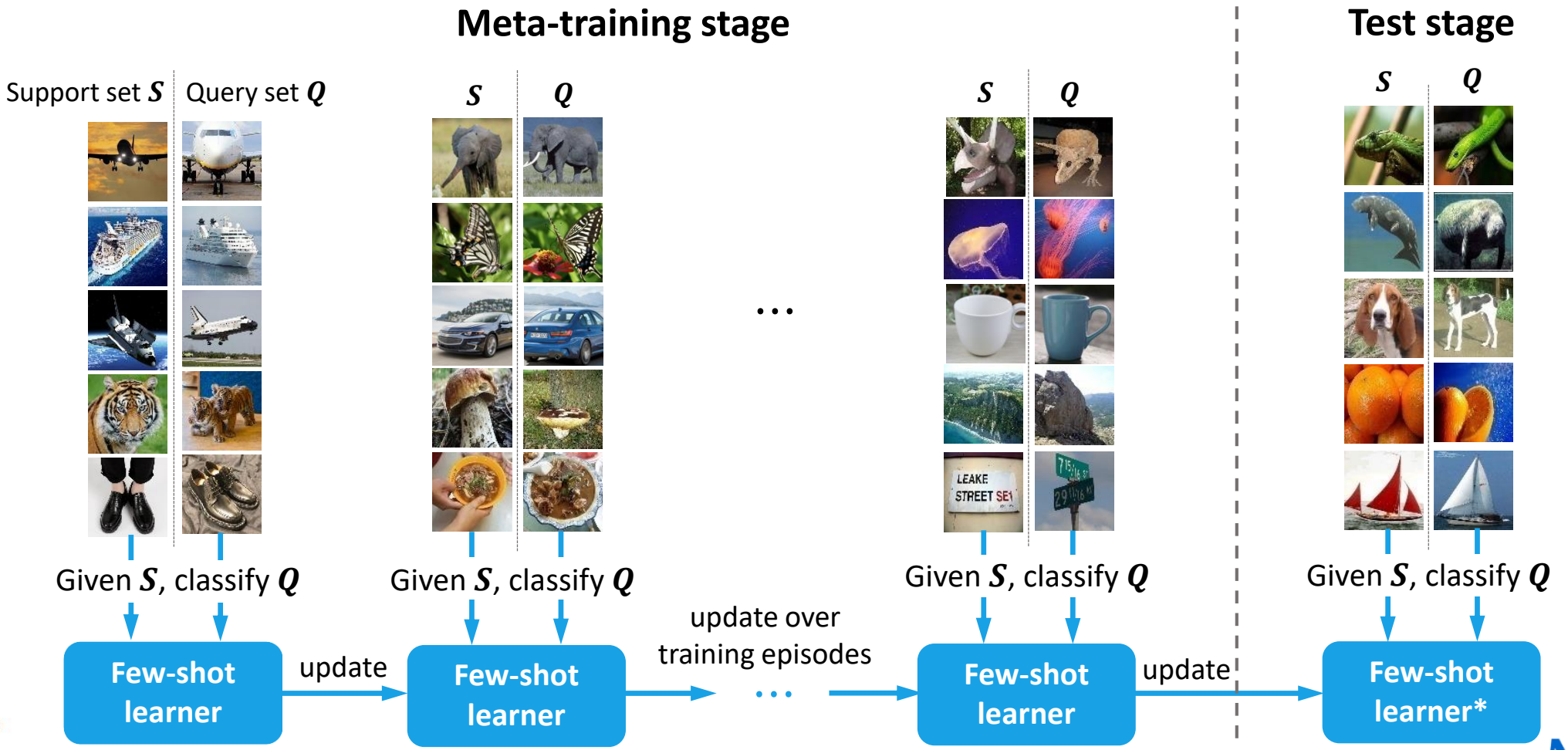
Jun 11 @ ICML 2019

KAIST School of Electrical Engineering

Korea Advanced Institute of Science and Technology

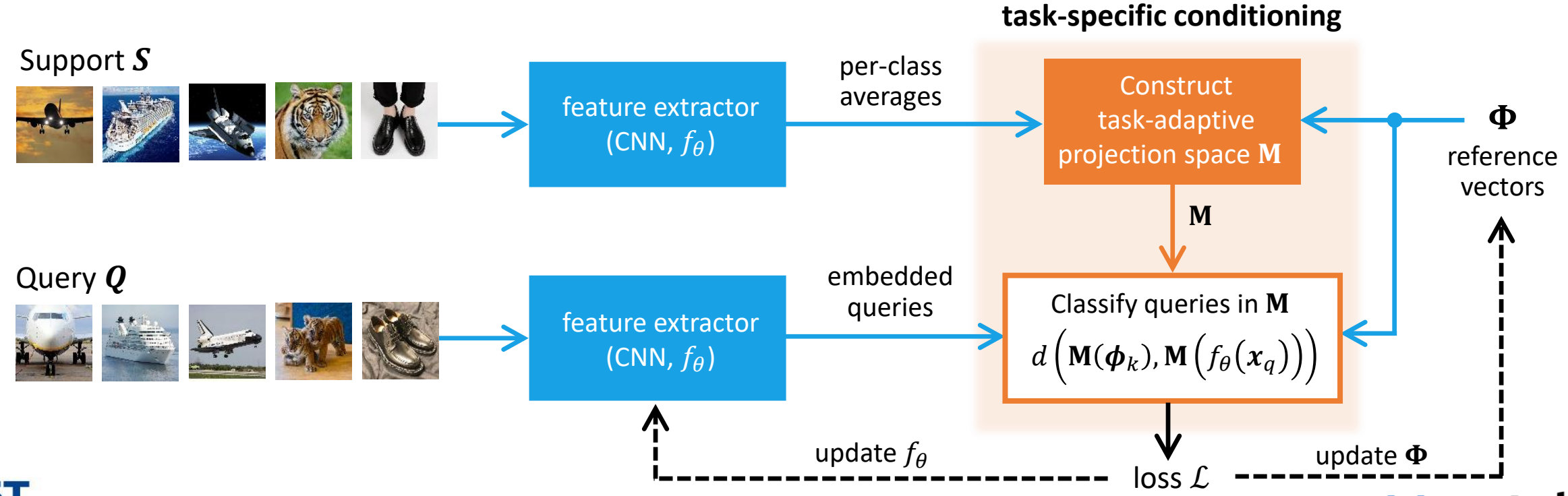
Few-Shot Learning

- Handling previously unseen classification tasks (episodes)
  - Training model with widely varying episodes (episodic training) [Vinyals et al., 2016]



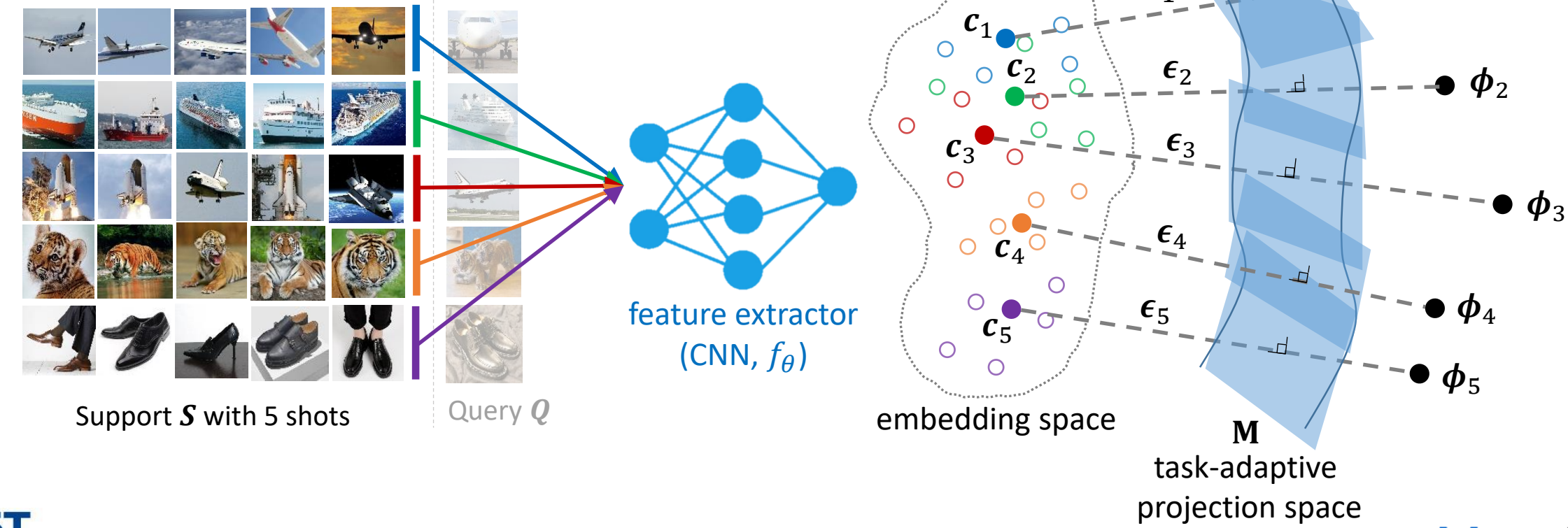
TapNet: Task-Adaptive Projection Network

- Model description (three key elements)
  - Feature extractor f_θ
  - Learnable reference vectors $\mathbf{\Phi} = [\boldsymbol{\phi}_1; \cdots; \boldsymbol{\phi}_N]$
  - Task-adaptive projection space $\mathbf{M}$
    - » Project references $\mathbf{\Phi}$ and embedded queries to $\mathbf{M}$, and apply metric-based classification



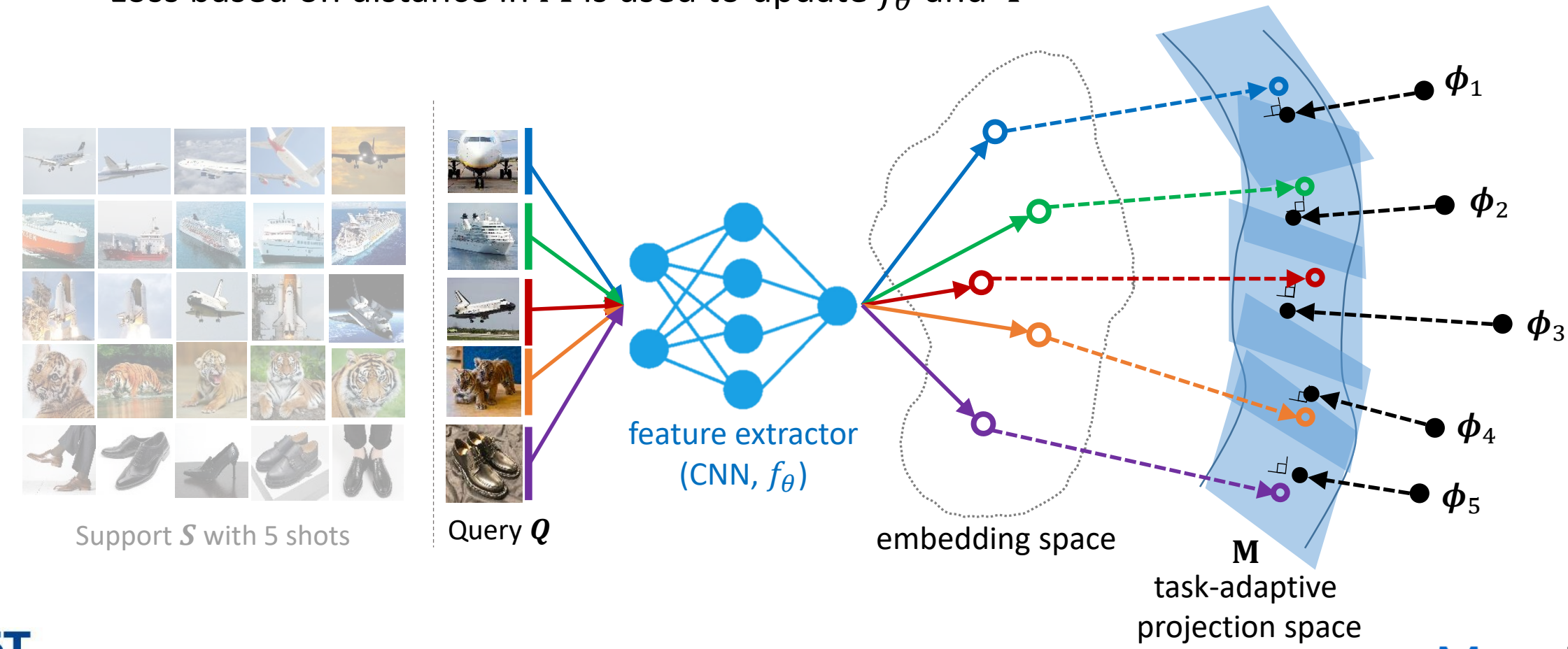
How to Construct Projection Space M

- Construction of projection space via linear nulling
  - Error vector between per-class average $\boldsymbol{c}_k$ and reference $\boldsymbol{\phi}_k$ should be zero-forced in $\mathbf{M}$.
  - For every episode, compute $\mathbf{M} = \mathrm{null}([\boldsymbol{\epsilon}_1; \boldsymbol{\epsilon}_2; \boldsymbol{\epsilon}_3; \boldsymbol{\epsilon}_4; \boldsymbol{\epsilon}_5])$ anew



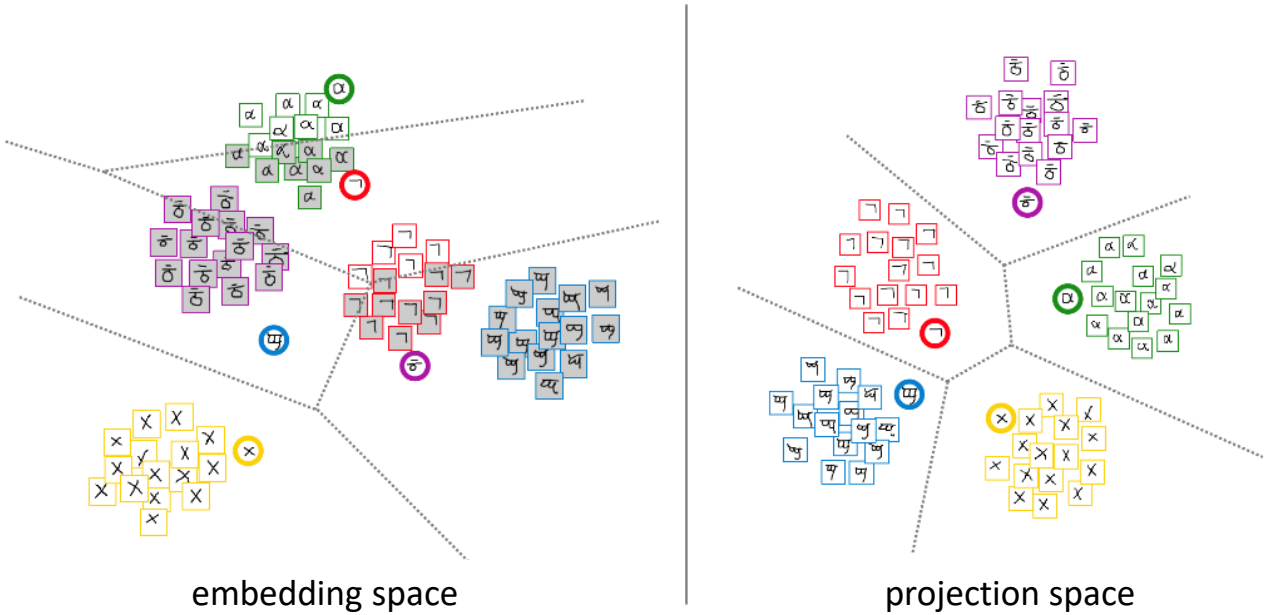
Classification and Learning

- Classifying in the task-adaptive space
  - Project $\boldsymbol{\Phi}$ and embedded queries to $\mathbf{M}$ → Classify the projected queries with projected $\boldsymbol{\Phi}$.
  - Loss based on distance in $\mathbf{M}$ is used to update f_θ and $\boldsymbol{\Phi}$



Observations

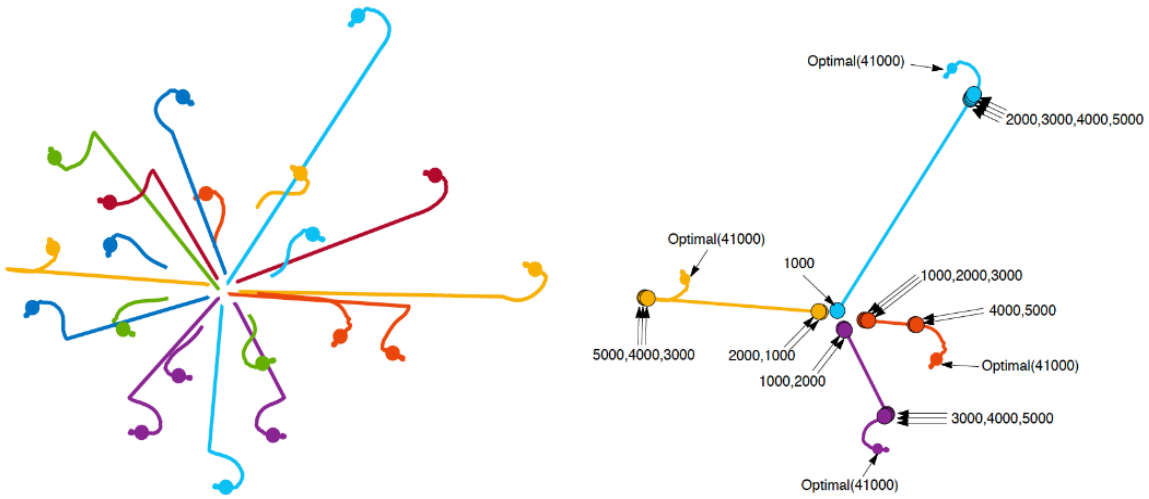
- Projection space gives better separation of classes
- Reference vector tips actually grow apart with training



embedding space

projection space

t-SNE plot



Learning trend of Φ

Results and Conclusions

Visit Poster #4 in Pacific Ballroom for further results!

- Non-learning-based and explicit task-adaptation method
- Excellent generalization performance

Methods	5-way *mini*ImageNet	
	1-shot	5-shot
Matching Nets (Vinyals et al., 2016)	43.56 ± 0.84%	55.31 ± 0.73%
MAML (Finn et al., 2017)	48.70 ± 1.84%	63.15 ± 0.91%
Prototypical Nets (Snell et al., 2017)	49.42 ± 0.78%	68.20 ± 0.66%
SNAIL (Mishra et al., 2017)	55.71 ± 0.99%	68.88 ± 0.92%
adaResNet (Munkhdalai et al., 2018)	56.88 ± 0.62%	71.94 ± 0.57%
Transductive Propagation Nets (Liu et al., 2018)	55.51 ± 0.86%	69.86 ± 0.65%
TADAM-α (Oreshkin et al., 2018)	56.8 ± 0.3%	75.7 ± 0.2%
TADAM-TC (Oreshkin et al., 2018)	58.5 ± 0.3%	**76.7 ± 0.3%**
TapNet (Ours)	**61.65 ± 0.15%**	**76.36 ± 0.10%**

Methods	5-way *tiered*ImageNet	
	1-shot	5-shot
MAML (as evaluated in (Liu et al., 2018))	51.67 ± 1.81%	70.30 ± 1.75%
Prototypical Nets (as evaluated in (Liu et al., 2018))	53.31 ± 0.89%	72.69 ± 0.74%
Relation Nets (as evaluated in (Liu et al., 2018))	54.48 ± 0.93%	71.31 ± 0.78%
Transductive Propagation Nets (Liu et al., 2018)	59.91 ± 0.94%	73.30 ± 0.75%
TapNet (Ours)	**63.08 ± 0.15%**	**80.26 ± 0.12%**